A Recovery and Comparison of Shortwave Radiances Measured by CERES Instruments Operating on TRMM and Terra Satellites

Z.Peter Szewczyk *
Science Applications International Corporation (SAIC)

Abstract

A comparison of shortwave unfiltered radiances measured by CERES instruments operating on two different platforms, TRMM and Terra satellites, is presented. A focus of this paper is twofold: a summary of the recovery of data obtained by the CERES instrument (PFM) failing sensors at the end of its useful life in April and June of 2000, and validation of the recovered data. The data recovery is necessary as deteriorating electronics of the PFM polluted data with cross-talk and noise preventing data processing. It is shown that a pattern recognition is an effective strategy in the data clean-up. The validation is performed by comparing shortwave measurements obtained by the CERES instrument (FM1), on board the Terra satellite, to the recovered PFM data. Comparisons are made for the data collected when orbits of both satellites cross, and viewing geometries of the instruments match.

Keywords: calibration, noise removal, pattern recognition, neural networks, orbital crossings

1 Introduction

A Clouds and Earth Radiant Energy System (CERES) instrument is a scanning thermistor bolometer designed to measure reflected solar radiation and outgoing longwave radiation from the Earth for radiation budget studies[1]. The scanner has three channels: total (0.3 -100 μm), shortwave (0.3 - 5 μm), and a longwave window (8 - 12 μm) which measure different part of a spectrum. The first CERES instrument, proto flight model (PFM), had been successfully operated on the Tropical Rainfall Measuring Mission (TRMM) satellite, before it got turned off in mid June of 2000 due to a failure of its electronics[2]. The PFM operation before its failure is particularly important from the research standpoint as it offers a unique opportunity to cross-validate its measurements with two additional CERES instruments, flight model 1 and 2 (FM1 and FM2), operating on board the Terra satellite since the beginning of 2000. While still in operation, PFM collects measurements in all three channels. However, these measurements have not been processed or validated, as the instrument exhibits problems with its electronics resulting in increased noise and cross-talk between radiometric channels and house-keeping (diagnostics) data. Therefore, in order to make radiances available for research, the data must be first recovered by removing the noise and cross-talk.

PFM measurements taken by its failing sensors can be made available for research studies provided processing removes the cross-talk and reduces noise content. Out of the three channels, processing shortwave channel data is particularly promising as there is enough information to describe the noise content in measured radiances. Since a shortwave channel measures a light component of the spectrum, nighttime data collected during one orbit can be used to characterize the noise, and then used on daytime data to reduce the noise content. Such an approach is not readily available for the other two channels. A cross-talk can be effectively removed by eliminating any value of the shortwave channel which coincides with other channels or hous-keeping data.

Filtering is commonly used to reduce a noise content in signals of interest, and a spectral analysis identifies frequency content to be removed. However, this approach is deemed inappropriate for processing

* *e-mail:* z.p.szewczyk@larc.nasa.gov; tel.(757) 825-7023; 1 Enterprise Pkwy, Hampton, VA 23666

Table 1: A CERES data record

space 1	earth	space 2	space 3	earth	space 4
1:40	41:280	281:330	331:380	381:620	621:660

the shortwave channel data, as daytime signals contain frequencies which coincide with frequencies found in the electronic noise. Therefore, a pattern recognition strategy is used instead to reduce the noise content in the data of interest. The electronic noise is characterized based on the nighttime data, and the noise signatures are produced. Then, the noise signatures are applied to daytime data to reduce the noise content. The scope of processing is quite large as each day of data contains about 13,000 records or vectors that correspond to one full sweep of the scanner, and that there are about four weeks of data that need to be cleaned and processed. Therefore, a feature-sensitive neural (FSN) network is utilized here as a pattern recognition technique as it can efficiently store and retrieve a large amount of information.

The recovered PFM shortwave data are validated by comparing them to data collected by the FM1 instrument on board the Terra satellite. The TRMM and Terra satellites are in two different inclination orbits, and therefore their ground tracks intersect. These orbital crossings offer a unique opportunity to compare radiances measured by the instruments provided the instrument viewing time and angles (azimuth and zenith) are matched within a prescribed tolerance. Once the matched data have been extracted, a statistical analysis can be performed on both data distributions. It is assumed that the distributions are statistically independent, and to reduce spatial noise, averaging is performed on a grid. Several different statistical measures are computed for the data.

2 PFM shortwave data recovery

The data recovery is basically a two-step procedure. In the first step, the cross-talk between a shortwave radiometric channel and other channels and house-keeping data is removed. In the second step, nighttime measurements of each orbit are used to obtain shortwave noise signatures, and a noise signature is subtracted from the the daytime measurements.

2.1 CERES radiometric data

A CERES instrument is a scanner whose a full sweep across the Earth lasts 6.6 sec. With a sampling rate of $100[sec^{-1}]$, one full sweep or record of data contains 660 samples. A full sweep consists of five major parts. It starts by looking into space on one side of a globe, scans across the Earth, then looks into space and internal calibration on the other side, scans across the Earth again, and finishes by looking into space again. Therefore, each data record reflects that sweeping motion and all 660 samples are divided up accordingly. The Table 1 provides the details of sample positions for the PFM.

A shortwave channel measures the light portion of radiation. Therefore, the nighttime records only show the electronic noise. In the daytime records, the noise is only shown during the looks into deep space, and the rest of a record contains vicarious or Earth viewing radiometric measurements. Each record of data belongs either to a nighttime or daytime portion of an orbit, and one day of data contains almost 15 orbits.

2.2 Cross-talk

A failing multiplexer caused several anomalies with data collection. The radiometric channels were spilling their values to the adjacent channels, the house-keeping (diaganostic) data were polluting radiance measurements, and also bit toggling resulted in random “spikes” in the data of interest. Since any sample could have been affected by the cross-talk, all samples of a record needed to be checked for any anomaly. If found, such a sample was simply removed from further processing.

Random bit toggling detection is quite challenging as there is no clear definition what it is, as any rapid value change within one sample looks like a spike in data. Spikes are easy to detect in the nighttime and space look data as the instrument is looking into the darkness. However, spikes also occur naturally in the healthy daytime data, but their distribution is quite different. Therefore, analyzing spike distribution in healthy data and forcing that statistical distribution on polluted data is an approach taken in this work.

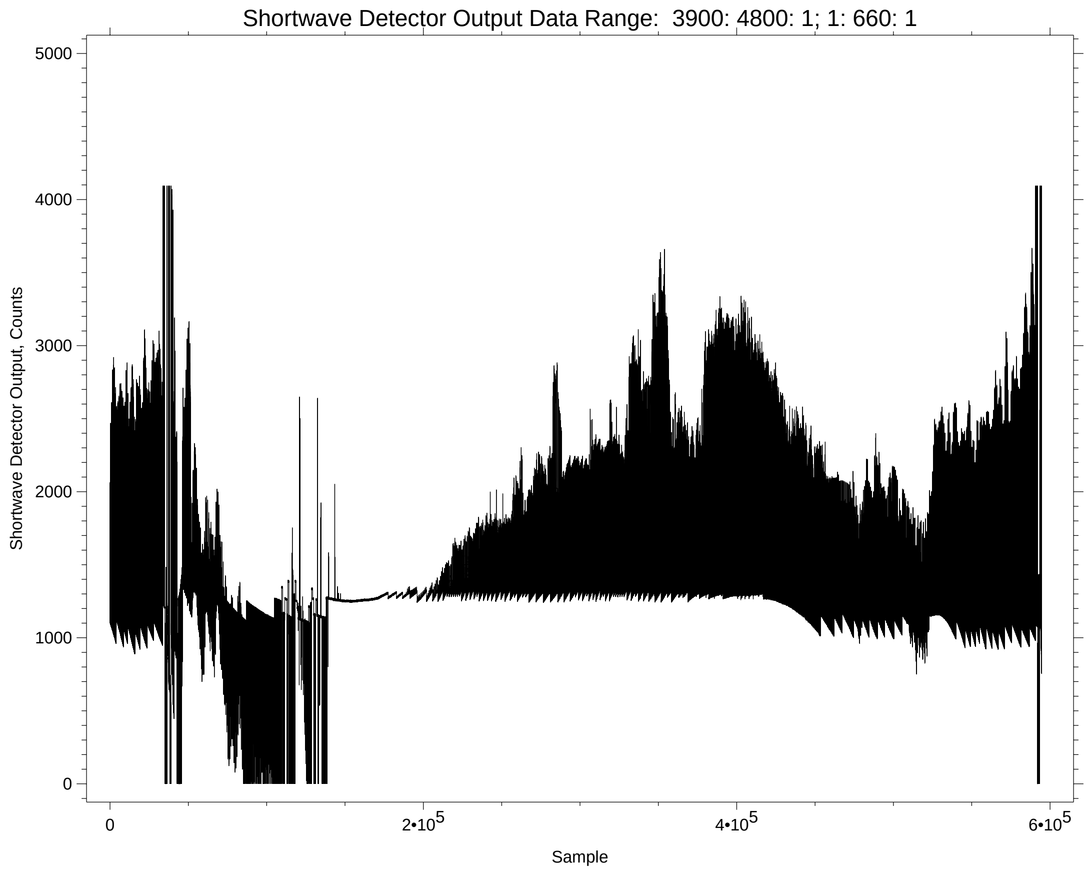

Figure 1: A sample orbit of shortwave measurements with noise and cross-talk

The data analysis indicated that less than 10% of data were affected by the cross-talk. However, its severity precluded radiance processing with the production code whose quality control parameters eliminated most of data records from processing. Therefore, off-line processing recovered almost 90% of valuable data.

2.3 Noise

A healthy instrument showed very small and stable electronic noise on an order of one count. That value started to go up in March of 2000 as a result of failing electronic converters and reached the value of 4 counts on average at the end of the instrument useful life. Data analysis showed that within one orbit the noise was quite stable, only slightly changing its amplitude and phase angle. A spectral analysis of noise signals showed several dominant low frequency components also found in the daytime data. Hence, filtering with a notch filter was considered to have been inappropriate for the application. Based on the assumption that the noise was stable within an orbit, the subtraction of a noise signature was used as the alternative approach.

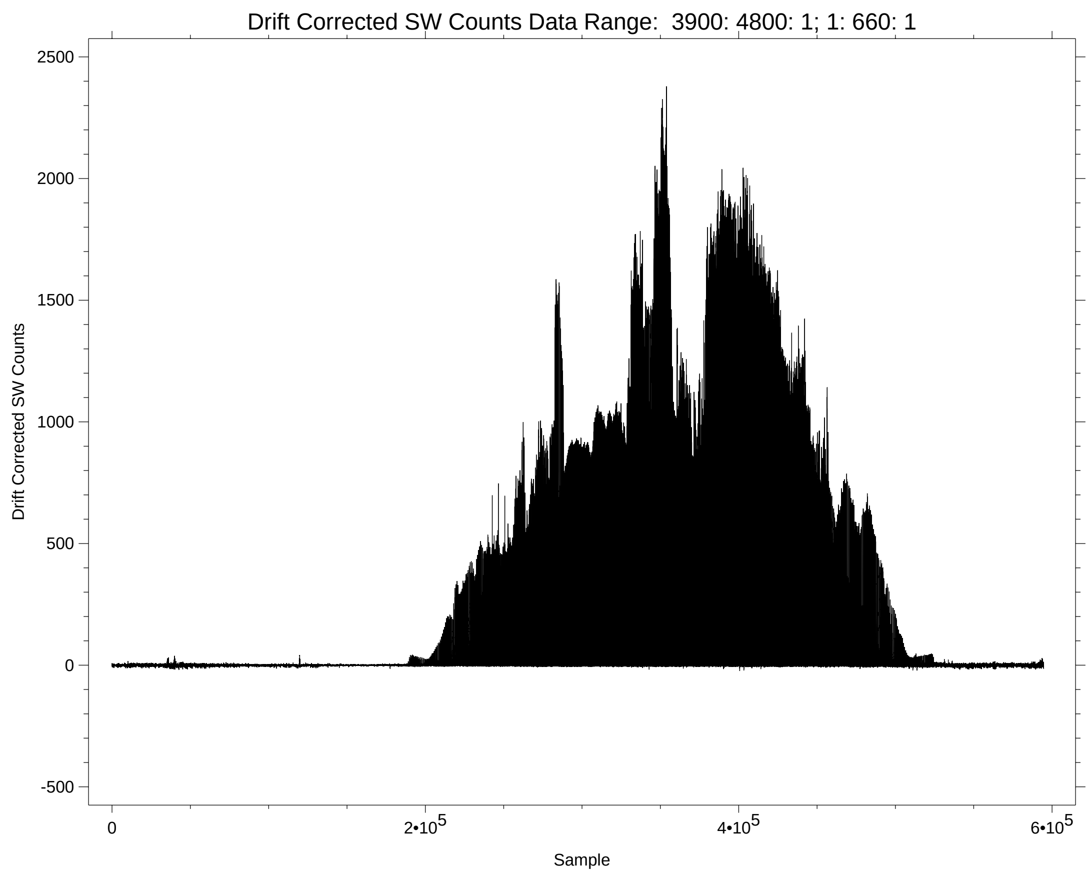

Figure 2: The same orbit after the pattern recognition based clean-up

In this application, a noise signature is a cluster examplar (see the Appendix for information on feature-sensitive neural (FSN) networks.) A set of noise signatures is obtained via the training of a FSN network on a collection of nighttime noise vectors for each orbit. The set is then used to reduce the noise content in the daytime data. Selection of a noise signature for a given daytime signal is based on matching the space look portion of a record (1 and 4). The closest match is subtracted from the signal of interest. The process of obtaining noise signatures is repeated for each orbit.

2.4 Procedure

Based on the discussion presented in the previous subsection, the following processing procedure is devised (all steps were repeated for each orbit of data processed):

1. Remove cross-talk caused by other radiometric channels or diagnostic data.
2. Remove random spikes from the daytime data.
3. Correct a data drift in a given record, and convert shortwave radiances.
4. Use the nighttime data to obtain a set of noise signatures.

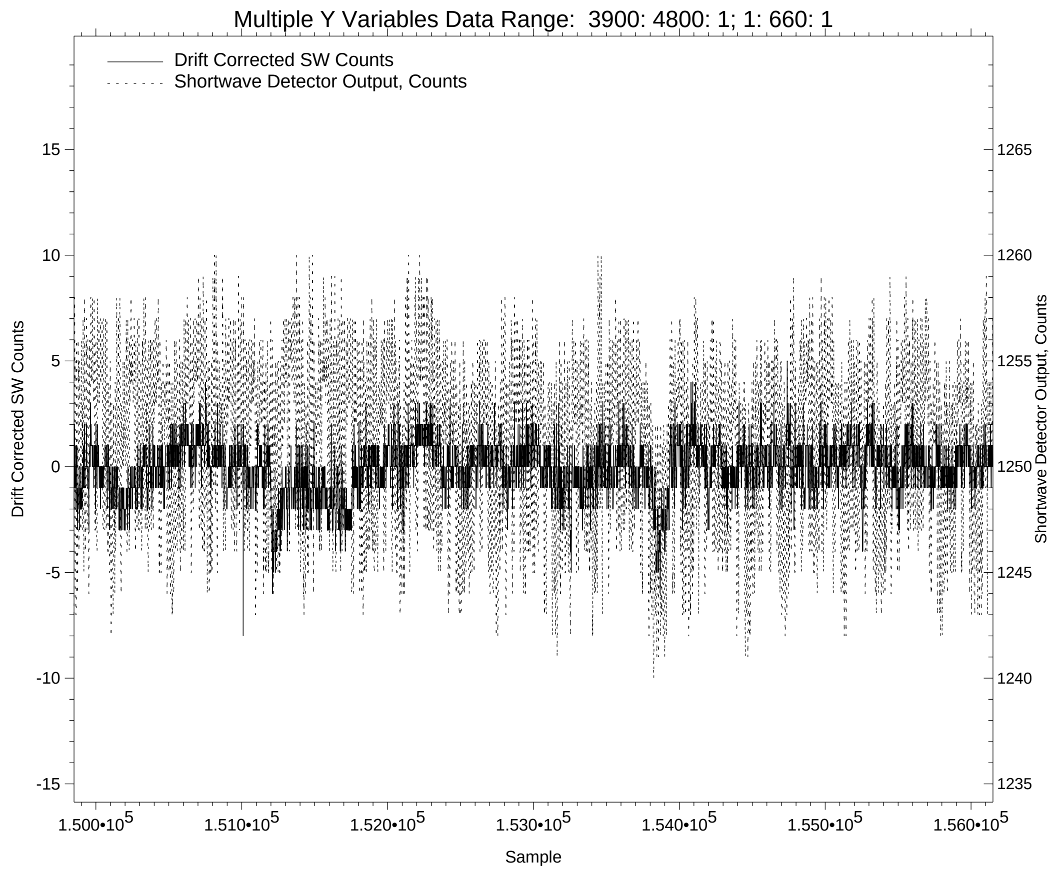

Figure 3: A nighttime data before and after the noise removal

5. Use space look portions of a daytime scan to identify the closest noise signature to be subtracted from the scan.

2.5 Results

Instrument measurements are recorded in counts whose dynamic range is 0 - 4095. Since instrument measurements drift within its dynamic range, there are updates performed automatically to adjust the position of zero within the range. When the instrument data are processed, the drift is corrected, and the measurements are converted into physical units where $1[count] = 6.05[Wm^{-2}sr^{-1}]$. Numerical values shown in plots and tables are in *counts* for consistency with the data originally measured by the instrument. An example of one orbit of polluted measurements taken by the CERES instrument on 04/09/2000 is shown in Figure 1. In Figure 2 the same orbit is shown after applying the drift correction and removing the cross-talk and noise. Figure 3 shows the nighttime data (noise) before, labeled as “Shortwave Detector Output Counts”, and after, labeled as “Drift Corrected SW Counts”, noise signals are processed by a FSN network. In this example, the average energy per sample is reduced from 9.06 $[count^2]$ to 1.33 $[count^2]$. Figure 4 shows a space viewing portion of a scan processed as a noise input to identify the noise signature. And finally, Figure 5 shows an earth viewing portion of daytime data. Figure 5 shows that spurious spikes in the measurements got removed, and the change to the signal itself is limited to what is shown in Figure 3.

Processing summary of the data are shown in Table 2. In the second column, the average percentage

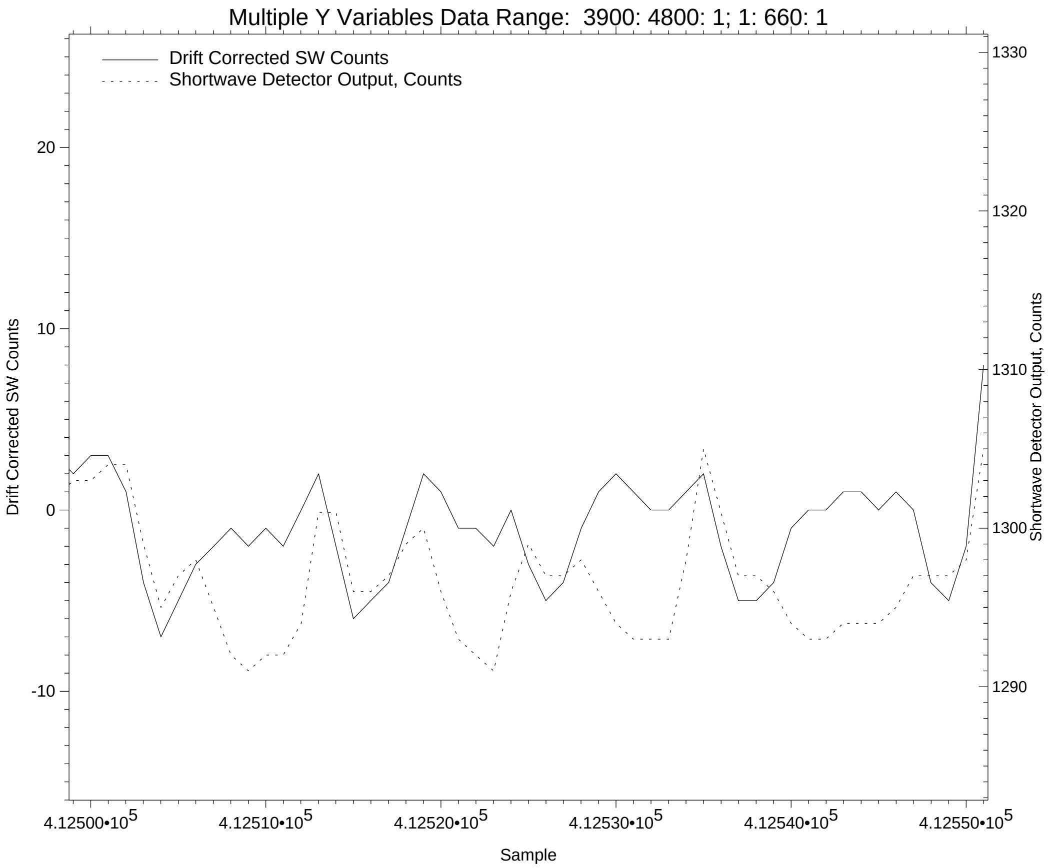

Figure 4: A space portion of a scan

of data lost due to cross-talk is shown. The next two columns show the nighttime (nE) and the daytime (dE) noise energy computed in $count^2$. The last two columns show the noise energy after the noise removing technique is applied.

3 Recovered data validation

The recovered PFM shortwave data quality is assessed by comparing them to data collected by the FM1 instrument on board the Terra satellite [3]. The next section describes the data collection process, followed by a brief description of the statistics applied to the data, and the results.

3.1 Selecting orbital crossings data

The TRMM and Terra satellites are in 35° and 98.2° inclination orbits, respectively. Their ground tracks cross 15 times a day. However, if we assume that for an orbital crossing to occur both satellites should be at the same location no more than 20 min. apart, the effective number of crossings per day can be as little as eight. Scanning planes orientation of both intruments are different, and their orientation with respect to the Sun is also different. The orientation depends on the scanning head orientation (a relative azimuth angle) and the satellite heading. In general, there may be very few scans which coincide for both instruments.

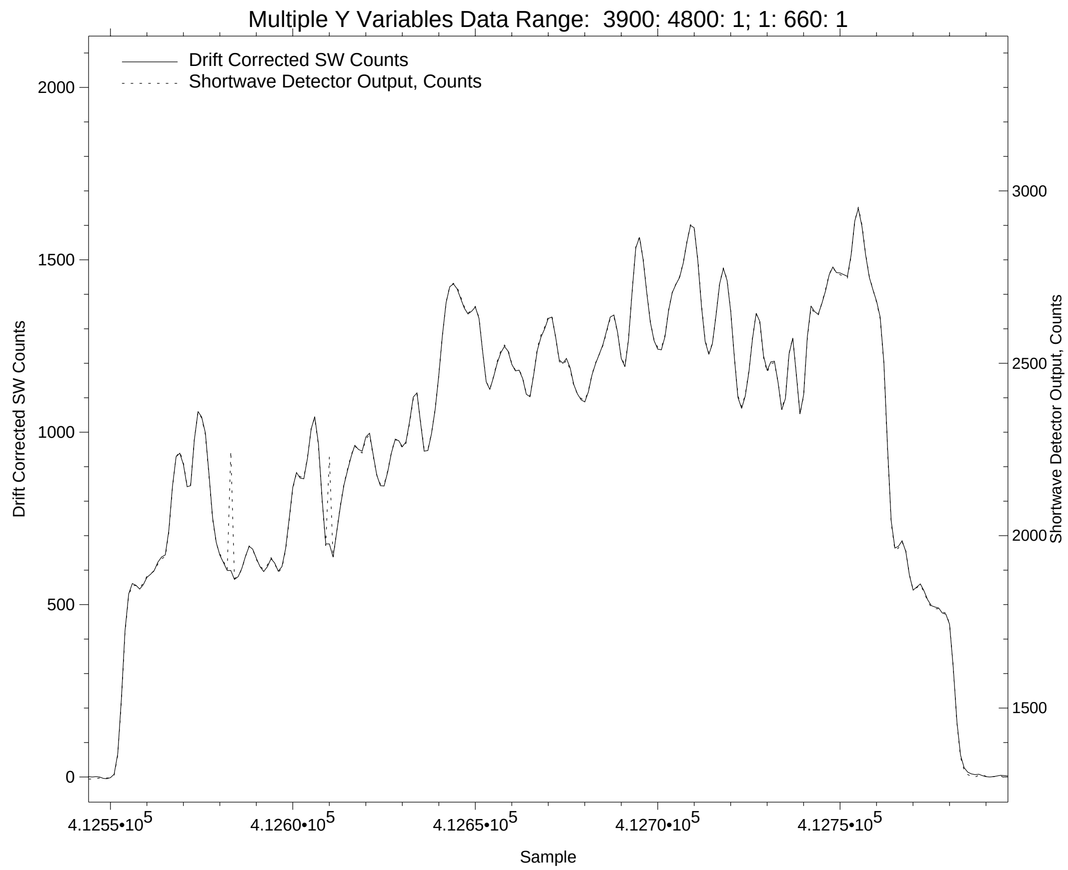

Figure 5: A daytime partion of a scan

However, if one of the instruments rotates while scanning, the scanning planes can coincide quite often, and produce measurements suitable for comparison. If the relative azimuth angle difference is less than $10°$, then their measurements are compared. Since the solar zenith angle is different for each instrument, shortwave measurements of one need to be normalized with respect to the other. CERES instrument scans across the Earth, and an elevation angle of the scanner determines the viewing zenith angle. For the shortwave data, it is required that the viewing zenith angle is within $10°$ from each other to produce a footprint to be considered ($0°$ is at the nadir.) The size of a footprint is different for both instruments because the satellites flying altitudes are different. A PFM footprint is 5km by 10km at the equator, and the FM1 is twice that size. Footprints are compared if there is at least 75% overlap between them. In order to reduce the spatial noise, averaging of shortwave radiances is performed on a grid of the size $1°$ latitude by $1°$ longitude assuming that at least 30 footprints are required for an average to be valid.

3.2 Statistics

Shortwave measurements of each satellite collected over a given period of time form two sets of values which can be analyzed for their statistical equivalence. A sample in each set is an average value of a "real" radiance measured by each instrument for a given grid. Since each grid has two averages associated with it, an analysis of their difference provides information about how consistent the measurements are. A standard statistical measure of such consistency is the α-confidence test. The test gives an interval for the mean value which will

Table 2: Processing summary of the shortwave radiance recovery

dates	xtalk	nE_{noise}	dE_{noise}	$nE_{no-noise}$	$dE_{no-noise}$
04/01-09/2000	8.2%	9.3	7.1	1.2	3.8
06/02-13/2000	10.5%	14.5	10.4	3.3	7.6

Table 3: Statistical data

dates	n-data	μ_{FM1}	μ_{PFM}	σ_{FM1}	σ_{PFM}	μ_{diff}	σ_{diff}
03/01-31/2000	762	75.4	75.3	54.6	55.3	0.08	5.8
04/01-09/2000	792	82.3	82.1	55.2	56.1	0.24	5.7
06/02-13/2000	331	85.6	85.5	64.7	65.5	0.13	4.8

contain any new average with the probablity of α, typically set to 95%. One would hope that this interval is smaller than 1%. Each set of averages contains the same amount of elements. They form two independent distributions, but with the values that are paired up (two instruments estimating the same radiation.) It is therefore reasonable to test their means to see whether there is a significant difference between them. The standard Student's $t-test$ is used in this work, and also the χ^2-test is used to check whether both distributions of averages are significantly different.

To summarize, the following standard statistical measures are computed for the collected orbital crossings data:

1. $\alpha-test$ with $alpha$ set to 95%,
2. Student's $t-test$ for significantly different means assuming paired-up values,
3. χ^2-test for significantly different distributions.

3.3 Results

In order to compare and validate the recovered PFM shortwave measurements, the reference for comparison is established based on data collected by both instruments in March, 2000. The same statistical measures are computed for three time periods, and results of all tests are summarized in Tables 3 and 4.

The value listed as an $\alpha-test$ determines the size of an interval for the means. For the March and April data, this interval is less than 1%. For the June data it grows to 2%, however, that increase should be attributed to the increased uncertainty due to fewer samples available. The $t-test$ also indicates that the means difference is not significant (≥ 0.05), and also the χ^2-test shows that it cannot be disproven that the distributions are the same.

Table 4: Statistical analysis of shortwave radiances

dates	$\alpha-test$	$t-test$	χ^2-test
03/01-31/2000	0.41	0.69	1.0
04/01-09/2000	0.40	0.11	1.0
06/02-13/2000	1.70	0.89	1.0

4 Closing Remarks

A pattern recognition approach to removing data pollution caused by cross-talk and electronic noise has merit. It is demonstrated that vector matching and noise subtracting avoids indiscriminate data changes offered by filtering techniques. A feature-sensitive neural network is an effective associate memory tool that can handle processing hundred of thousands of records of data as required by the application.

Several standard statistical measures have been computed to compare and validate the recovered data. They seem to indicate that measurements of FM1 and PFM belong to the same distribution. The $\alpha - test$ shows that there is the 95% likelihood that radiances measured by each instrument are within 1% for the March (healthy instrument) and April data, but that value grows to 2% for the June due to fewer samples available.

It is important to underscore the fact that 21 days of recovered shortwave radiances along with the data collected in March of 2000 constitute bases for establishing the consistency in long-term Earth energy budget measurements.

5 Acknowledgements

This work was supported by the Earth Science Enterprise of NASA Langley Research Center under contract NAS1-19570 to Science Applications International Corporation (SAIC.) The author would like to thank Dr. J.Kibler for his encouragement to complete the data recovery work, and Dr. M.Haefflin for sharing his expertise in the calibration field.

References

[1] B.A. Wielicki et al. Clouds and the earth's radiant energy system (ceres): an earth observing system experiment. *Bul. Amer. Met. Soc.*, 77:853–868, 1996.

[2] J.J. Chapman et al. Assessment of selected ceres electronic component survivability under simulated over-voltage conditions. *Proc. Soc. Photo-Opt. Instr. Eng.*, 3750, 1999.

[3] M. Haefflin et al. Inter-calibration of ceres and scarab earth radiation budget datasets using temporally and spatially collocated radiance measurements. *Geophysical Research Letters*, 28(1):167–170, 2001.

[4] Z. P. Szewczyk and P. Hajela. Feature-sensitive neural networks in structural response estimation. In *Intelligent Engineering Systems Through Neural Networks*, pages 159–164. ASME Press, 1992.

6 Appendix

There are several different applications for feature-sensitive neural networks which include direct and inverse function approximation, associative memory recollection, classification, and also pattern recognition. These networks may be thought of as generalized look-up tables in which storing information is based on self-organizing clustering. Retrieving information from the network uses a vector matching and interpolating procedure. Since the data processing with FSN networks uses differences and averaging applied to each vector component independently, the technique is very efficient and capable of handling high-dimensionality inputs.

Feature-sensitive neural networks combine several techniques to become a useful tool for a pattern recognition type of problems [4]. They incorporate a clustering routine with a Δ-rule to yield a near optimal distribution of cluster examplars. In a more general setting, each cluster examplar is associated with an output vector (a cluster label) to enable approximating a system response for any new input. In the present application, a cluster examplar is used as a noise signature, therefore further genaralization is not needed.

In order to show how a FSN network works, we assume that there is a set, S, of the size, M, of discrete data forming a database or a table which relate n input quantities to m output quantities representative of some mapping $\phi : \mathbf{x} \in R^n \mapsto \mathbf{y} \in R^m$. Therefore a database entry or a pattern, p_j, is simply a row of data formed by juxtaposition of both input and output quantities, thus $p_j \equiv [x_1, \ldots, x_n, y_1, \ldots, y_m]_j, \quad j = 1, \ldots, M$. Weight vectors of the feature-sensitive neurons are simply defined as

$$\mathbf{w}_i \equiv [w_1, \ldots, w_n, w_{n+1}, \ldots, w_{n+m}]^\top \in R^{n+m},$$

and they belong to the same space as entries in a database. Since a clustering routine can operate on virtually any part of a weight, a FSN network is very flexible in approximating various relations.

In order to train a FSN network, i.e. find a near optimal location of cluster examplars and associate them with labels, we assume that the initialization part of the ordering routine is executed with a selected resolution δ for all $n+m$ components of a pattern. Just one pass of all patterns results in A cluster examplars or weights, w_α. After initialization, the delta rule

$$w^{new} = (1 - \theta(t_e))w^{old} + \theta(t_e)p_i, \quad i = 1, \ldots, M$$

is used in the clustering algorithm to further refine seperation between clusters, location of cluster examplars, and the value of cluster labels. Here, a new entry is labeled p_i and θ is a small *learning* parameter which value is function of training epochs, t_e, and it monotonically decreases to zero. It is important to note that this algorithm is very efficient in defining a network and never fails to train. However, its main utility consists in removing redundancies in a database and extracting essential information about behavior of a system rather than storing detailed information about the system.

In retrieving information from the network, a matching procedure is used. Assume that some input vector, $\hat{b} \equiv [\mathbf{x}, *]^\top \in \mathbf{R}^{n+m}$, is presented to the network. By setting a part of this vector to a NaN value, the search routine matches its $\mathbf{x}$ components with corresponding components of weights in the network. Thus,

$$\forall \alpha \in A \quad \delta_\alpha = \sum_{l=1}^{n} \|p_l - w_{l\alpha}\|$$

and then, the closest match can be identified,

$$\forall \alpha \in A \quad \exists \alpha^\star$$

$$\delta_{\alpha^\star} = \sum_{l=1}^{n} \|p_l - w_{l\alpha^\star}\| \leq \delta_\alpha.$$

to be returned as the network response.